



**NOAA**  
**FISHERIES**  
**SEFSC**

# Overview of Fisheries Dependent Data Quality and Data Management

SEFSC Program Review

June 5, 2013

# Presentation outline



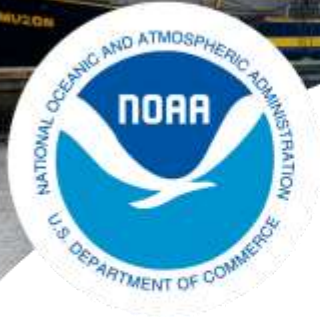
**NOAA**  
**FISHERIES**  
**SEFSC**

What is quality?

SEFSC vision

SE data systems: quality and management

SEFSC current and near-term projects and expected benefits



**NOAA**  
**FISHERIES**  
**SEFSC**

What is quality?

- ❖ Defect free product that meets the needs of the user
  - **Design to reduce errors**
    - Standardized data collection protocols
    - Understandable instructions for self-reported data
    - Units familiar to data collector/provider
  - **Auditing and editing**
    - Validate formatting
    - Out of range checks
    - Cross check entry vs other data collections
    - Correction of entry



**NOAA  
FISHERIES**

**SEFSC**

What is quality?

- ❖ Defect free product that meets the needs of the user
  - **Meet user needs**
    - Complete metadata describing information collected
    - Adequate to meet data needs of user
    - Short time before information is available
      - We can control for information collected directly by NMFS
      - Can't control for self reported data
    - Availability-accessibility for users
    - User feedback to better meet needs



**NOAA  
FISHERIES**

**SEFSC**

One clean copy of each data set

Data changes are tracked

Warehouse with all data from a single trip linked together

- **Logbook**
- **IFQ**
- **Trip ticket**
- **Dockside sampling**
- **Age and reproductive data**
- **Observer**

Multiple data for same trip are reconciled to identify errors and gaps

Reconciliation produces link for data loaded to warehouse by trip

User access to data provides 90% of requests for data



**NOAA  
FISHERIES**

**SEFSC**

## Coastal logbook

### Strengths

- Census of effort and catch from permitted vessels
- Species composition more accurate than dealer reports
- Regularly check for invalid data and multiple submissions
- 90% of trip report data available 55 days after receipt by SEFSC
- 90% of trip report data available 176 days after the end of the trip

### Weaknesses

- Discard reporting is unreliable
- Late submissions are common

### Solutions

- Reconcile with observer data to validate discard rate
- Contact vessel during the year to obtain delinquent reports
- Electronic logbooks



**NOAA  
FISHERIES**

**SEFSC**

## **HMS logbook**

### **Strengths**

- Large number of personnel reviewing logbooks relative to fleet size
- Regularly contact delinquent vessels
- Effort/catch information is at the set level

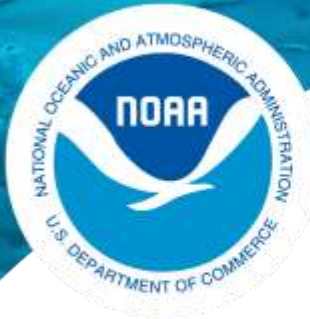
### **Weaknesses**

- Timeliness is issue – data are not complete until about March of the following year-also delays during the year resulting from waiting for all three reports (trip summary, set form, weigh-out) before data are made available
- Catch is in numbers, not pounds-must estimate pounds from weigh-outs

### **Solutions**

- Implement e-logbook
- Require logbook schedule number on trip tickets





**NOAA**  
**FISHERIES**  
**SEFSC**

## Puerto Rico trip tickets

### Strengths

- Only source of catch/effort information in PR
- Entered by PRDNR staff, so allows for some visual QC of data

### Weaknesses

- Lack of connectivity and slow internet connections have slowed current entry
- Some aggregate species groups in historical data
- Under/non-reporting by fishermen (as high as 50% non-reporting)

### Solutions

- Resolve connectivity issue
- Work with territory and fishermen to report species
- Work to identify non-reporters and increase enforcement presence
- Dockside validation of landings





**NOAA**  
**FISHERIES**

**SEFSC**

## USVI commercial catch reports

### Strengths

- Only source of catch/effort information in USVI
- Entered by USVI staff, so allows for some visual QC of data
- Yearly meeting with fishermen to explain how to fill out the catch form

### Weaknesses

- Lack of connectivity and slow internet connections have slowed current entry
- Suspected under/non-reporting by fishermen

### Solutions

- Address connectivity
- Dockside validation of landings



**NOAA  
FISHERIES**

**SEFSC**

## Commercial Fishery Observer Program

### Strengths

- Full debriefing of each observer by coordinators after trip
- Data collection standards developed by each program
- Data entry by each program, data proofing for key punch errors

### Weaknesses

- Not all programs have “range error checking” for data entry
- Each program developed observer data forms
- Some programs use Oracle and others use Access

### Solutions

- Develop “range error checking” software for each program
- Develop electronic offshore data entry
- Develop standard data format to warehouse data



**NOAA**  
**FISHERIES**

**SEFSC**

## Atlantic Trip Ticket

### Strengths

- Warehouse contains audited trip ticket data from each state on the Atlantic coast
- Data collection standards developed by partners
- Supports electronic reporting of trip ticket data
- Changes in electronic reports are tracked

### Weaknesses

- Warehouse changes not tracked
- Some partners not adequately supporting permits information

### Solutions

- Implement change tracking in warehouse
- Work with partners to provide staff time to keep permit data up to date



**NOAA  
FISHERIES**

**SEFSC**

## Gulf Trip Ticket

### Strengths

- Warehouse contains audited trip ticket data from each state on the Gulf coast

### Weaknesses

- Warehouse changes not tracked
  - Results in two versions of the same data
- Data collection standards for partners less extensive than Atlantic
- No standard process for making corrections
- Incomplete permit data

### Solutions

- Implement change tracking in warehouse
- Work to ensure that partners comply with standards
- All partners agree to process for making corrections

# Electronic commercial landings data



**NOAA  
FISHERIES**

**SEFSC**

## SAFIS/Bluefin data

### Strengths

- SAFIS is single online application, so any changes are immediately available
- SAFIS is able to handle entry for multiple partners and store data in the same structure
- Bluefin Data is desktop application that can be used with a slow connection

### Weaknesses

- Several states don't maintain permit data in SAFIS, which leads to validation exceptions and inability to load NC and FL data to SAFIS transactional table
- Bluefin data does not use a single structure to store data, but files with different formats and code systems-results in inability to reliably standardize to a single structure for data storage and access

### Solutions

- Implement standards for all trip ticket programs
- Require maintenance of permit tables from all partners
- Require electronic data be standardized to a single structure, so all partners use the same source data



**NOAA  
FISHERIES**

**SEFSC**

## **MRIP/MRFSS**

### **Strengths**

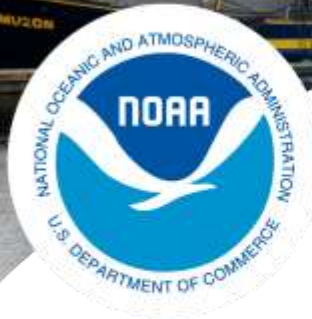
- Only time series of charter and private catch/effort data from NC-LA: starting in 1981
- Discards estimated throughout time series
- Recent improvements to survey design

### **Weaknesses**

- Changes in methodology over time and in the future
- Species ID for discards and unseen landings may be inaccurate
- High PSE for rarely encountered species



# Texas Parks and Wildlife Division recreational catch/effort



**NOAA  
FISHERIES**

**SEFSC**

## Texas Marine Sport-Harvest Monitoring Program

### Strengths

- Only time series of charter and private catch/effort data in TX
- Long time series, began in 1983

### Weaknesses

- Estimates are calculated for high-use and low-use seasons... not monthly or semi-monthly
- Estimates not made for all species
- Estimates in numbers only-we must estimate weight
- No discards recorded
- No shore mode

### Solutions

- Work with state to change protocol to collect variables needed for assessments





**NOAA  
FISHERIES**

**SEFSC**

## Southeast Region Headboat Survey

### Strengths

- Long time series began in 1972 and designed as census of federally permitted headboats NC-TX
- Electronic trip reports include: default settings, set parameters, range limits and required fields
- Dockside validation of effort
- Port agent review of reports for accuracy

### Weaknesses

- Limited validation of landings and discards
- Must estimate discards at each SEDAR-no accepted methodology used by program
- Late reporting

### Solutions

- On-board observers to validate discards
- Electronic logbook reporting begun this year
- Develop accepted method to estimate discards
- Increased enforcement of reporting regulations



**NOAA  
FISHERIES**

**SEFSC**

## Domestic Longline System (DLS)

### Strengths

- Several people conducting QC on weigh-outs
- SEFSC staff enter data-second opportunity for QC

### Weaknesses

- Small number of vessels must submit-HMS only
- No standardized form, so data is in a number of different formats

### Solutions

- Use e-reports from dealers



**NOAA  
FISHERIES**

**SEFSC**

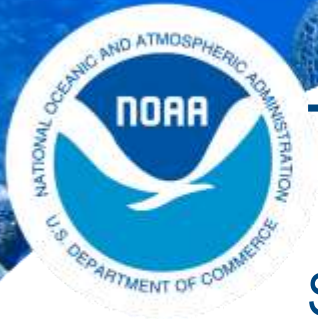
## Trip Interview Program

### Strengths

- Long time series of length composition data for key species in the SE
- Continental, good spatial coverage
- Validation of data during entry with current application
- High retention rate of samplers reduce some potential errors (e.g. species identification)

### Weaknesses

- Historic data contains values out of range
- Methodology not consistent over time or between samplers
- Purpose of collection changes over time
- Partners (Federal/Territorial/State) have a variety of sampling and data management approaches making full control of quality issues challenging



## **Trip Interview Program (continued)**

### **Solutions**

- Standardize methodology (currently underway)
- Develop standardized training curriculum and establish a sampling manual update schedule (currently underway)
- Devote staff time to ensure sampling staff adhere to standards (planned for next FY)
- Devote staff time to correct historic data (planned for next FY)
- Refine sampling design



**NOAA  
FISHERIES**

**SEFSC**

## Recreational/Commercial

### Strengths

- Primary source of length and age samples from private and charter boats in FL-LA
- Additional commercial samples to supplement TIP

### Weaknesses and threats

- GulfFIN funding priority for base commercial and recreational landings programs could result in substantial reduction in recreational length and age sampling
- Commercial sampling methodology differs from TIP

### Solutions

- Standardize commercial sampling with TIP

# Texas Parks and Wildlife Division recreational biosampling



**NOAA  
FISHERIES**

**SEFSC**

## Texas Marine Sport-Harvest Monitoring Program

### Strengths

- Only source of length data from recreational party/private trips
- Long time series, began in 1983

### Weaknesses

- Doesn't record weight of intercepted fish
- No samples from shore-mode

### Solutions

- Work with state to change protocol to collect variables needed for assessments





**NOAA  
FISHERIES**

**SEFSC**

## Southeast Region Headboat Survey

### Strengths

- Randomized trip sampling
- Main source of age samples in S Atlantic recreational fishery
- Able to collect reproductive samples as needed (fish not gutted)
- Systematic error checks

### Weaknesses

- No discard biosampling

### Solutions

- At-sea sampling to record discards on headboat and charters across entire region (ACCSP supports this)



# Panama City Laboratory



**NOAA**  
**FISHERIES**

**SEFSC**

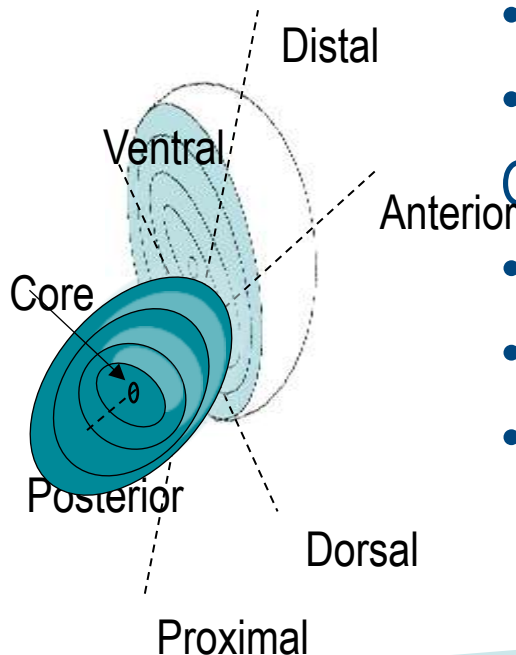
QA/QC

Otoliths:

- Processing standards based on species(relative size of otolith) isomet, hilquist, or benetec.
- APE's Percent error > 10% Typically around 4-5%
- Annual Otolith Standardization Meeting with GSMFC
- Reference sets shared, training slides(.ppt)

Gonads:

- Processing Standards
- Histology second readings
- Counts....developing imaging software



June 2013



**NOAA**  
**FISHERIES**

**SEFSC**  
**Panama City**  
**Laboratory**

# Data/Data Processing

## QA/QC Strengths

- TIP data no longer entered and proofed against the original field data sheet
- Implementation of proofing by Port Agents has significantly reduced errors.
- All biological data produced (ages, otolith weights, reproductive stages) are entered /proofed
- Data entries are then **proofed** against handwritten data sheets by another person
- Database safeguards (e.g., pull down data entry menus) have been added to reduce data entry errors
- FWRI, Headboat, Observer data, imported not entered
- All datasets submitted to SEDAR are preserved for future access

## Biological Sampling Database (Major improvement)

- Integration of TIP samples and the PC Lab data in one location
- Beginning in 2011, we no longer enter or proof TIP data since this data is now in BSD. However, we do continue to proof ages and otolith weights for TIP samples in BSD.

## Weaknesses to improve on

- Data from other sources which we do not control
- Data Standardization of data fields to assessors
- Errors/edits in data received
- Constrained by Access Database
- GSMFC data standardization issues (eg: max length/standard total Length)
- Reconciliation of data from outside sources (needs proofing component)
- No in house data manager

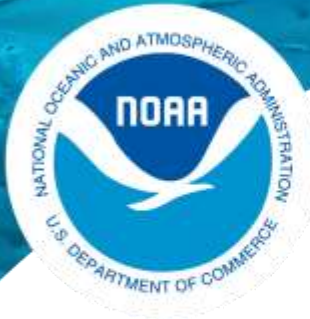


**NOAA**  
**FISHERIES**

**SEFSC**  
**Panama City**  
**Laboratory**

# Highest Needs for improvement:

- Data Quality and Access
  - Edits/Error Rates from Multiple sources for data that links to samples needs work
  - Migration to a centralized database system
  - Migration of Historical Bioprofiles data into BSD
  - Limitations on ability to standardize data fields from alternative sources
- Data Timeliness/Collections
  - Is slowly improving via communications with data suppliers
  - Increase sample collections from observer programs
  - Standardization of sample protocols and logistics(barcoding)
  - Increase recreational sampling component



**NOAA  
FISHERIES**

**SEFSC**

## SEFSC information resources

### Information Resources Management

#### Strengths

- Infra-structure sufficient to handle substantial increase in load
- Contractor support for developing systems

#### Vulnerabilities

- Partners not willing to accept new responsibilities for maintaining and interconnecting their data
- Insufficient staff to maintain complex systems, beginning to rely on contractors for tasks which require long term oversight
- Vulnerable network infrastructure - annual down times of weeks



**NOAA  
FISHERIES**

**SEFSC**

## Factors constraining FSD data management and quality

### Assessments

- Available analysts are tied up with SEDAR or ICCAT tasks
- Only 5 members of FSD with skills to participate in most data workshops

### Data Requests

- Increasing amount of data requests from SERO and Councils take remaining time from those with analytical capabilities

### Data Management

- Large number of data management tasks with only a few staff that are familiar with the databases

### Novel Skill

- Most staff are data processors-need additional training to develop quality procedures and programming capacity
- Have added contractors to assist with processing and reconciliation so that others can progress toward advanced data management

## Current SEFSC data quality and data management projects



**NOAA**  
**FISHERIES**

**SEFSC**

### Reconciliation of trip data

- Link trip data across multiple datasets
- Identify differences in common variables and make corrections

### Standardize look-up tables

- Standardize to a single code standard for gear, area, species, etc.
- Utilizing FIN standards to move to codes that all partners are currently using

### Probability based error checking

- Develop application for identifying possible outliers and flagging the value for further investigation





## Current SEFSC data quality and data management projects

### Biosample database (BSD)

- Biosample inventory application that utilizes the TIP database for interview information
- Reduces differences in the age and length data

### TIP Standardizing

- Updating collection methods, manual and training curriculum
- Planning review of TIP data to facilitate corrections

### Commercial Landings System (CLM)

- SEFSC system for handling electronic trip ticket data
- Error checking routines built to handle inconsistent formats/structure in files sent from E-reporting application vendor
- ACL reporting from this data built in APEX





**NOAA  
FISHERIES**

**SEFSC**

## Recent SEFSC data quality and data management projects

### Caribbean Commercial Landings System (CCL)

- First species specific reporting for the VI
- Online system for entering fisherman catch reports from PR and USVI
- Adds validation at entry, which was missing or incomplete for older data

### Vessel Monitoring System (VMS)

- Developing methods for identifying trips
- Data will be used to identify when trips have been made, but no logbooks submitted
- Also will be used to select trips for length composition sampling

## Recent/Current SEFSC data quality and data management projects



**NOAA  
FISHERIES**

**SEFSC**

### Trip Ticket Review System

- Port agents review electronic trip ticket data and make correction
- Suggested corrections are submitted to state programs

### Southeast Fisheries One-Stop-Shop (SEFOSS)

- Data access tool that will allow access to all data within the SEFSC data warehouse
- SEFSC will control treatment of the data, so it is consistent with the best available science



**NOAA  
FISHERIES**

**SEFSC**

## Expected benefits from current FSD projects

Reconciliation, probability based error checking, Trip ticket  
Review

- Expect that these projects will result in processes to identify errors, so available staff time can be used correcting errors, not trying to identify errors

TIP standardization, VMS

- Expect improvement in quality of length and age data
- Should result in less time spent by analysts trying to remove outliers from raw data.
- Increase in quality of trip selection

SEFOSS

- Should result in less time spent extracting data
- This will allow us to devote staff time to more quality control





# End

